University of Colorado Boulder

Predicting Occurrence and Duration of Flight Delays Due to Weather Conditions Using a Data-Driven Approach

Student :
Ella Bronaugh
Mayur Dalvi
Reza Naiman
Tristan Levy-Park

Teacher :
Dr. Kris Pruitt

Domain Experts :
Dr. Edward Ochoa
Leon Shen
Logan Gage

October 11, 2024

Contents

1 Introduction

Flight delays due to adverse weather conditions pose significant challenges to the aviation industry. Researchers in [3] found that extreme weather events were responsible for 32.6% of the total delay minutes recorded in the National Airspace System (NAS) from 2003 to 2015, with severe weather causing up to 82% of delay minutes in some instances. Such delays impact not only economic performance [6] but also environmental [7], [8] and social aspects. The annual economic impact of airline delays was estimated by one study to be $31.2 billion in 2010 and $40.2 billion in other estimates [1]. In addition, climate change and air transportation have a reciprocal relationship: aircraft emissions contribute to anthropogenic climate change while atmospheric changes directly impact operations in the airline industry [5]. Not only this–delays due to inclement weather are increasing over time. In comparing the impact of a winter storm in December 2021 versus December 2022, aggregated total passenger "dwell time" in airports saw an increase of approximately 12 million hours [9]. There are several factors contributing to increasing frequency of weather delays and the economic and environmental impacts are substantial; thus, the ability to accurately predict the occurrence and duration of flight delays and manage these disruptions is increasingly crucial. This study aims to answer the following question: are we able to improve upon existing models in order to classify weather-related flight delays and accurately predict the duration of these delays?

Previous studies, such as [2], [3], and [4], have highlighted the interaction between meteorological conditions and aviation operations, employing various machine learning and data-driven approaches to forecast delays. These findings emphasized the importance of understanding weather patterns and airport-specific vulnerabilities to optimize flight schedules and improve operational efficiency. One group of researchers in [4] examined arrival delays across the United States airport network using a variety of models. Their analysis, based on 2017 flight and weather data, found the GBM model to be the most effective. Although their study covered a larger geographic area, it mainly focused on classifying delays and faced challenges with imbalanced data and limited real-time data use. In a study conducted in 2024, a group of researchers in [2] expanded the scope by applying a wider suite of ML models to predict departure delays at three major international airports. Their models achieved high predictive accuracy, with rates of 0.749 for Incheon (ICN), 0.852 for John F. Kennedy (JFK), and 0.785 for Chicago Midway (MDW) in 2-hour forecasts. Although their study demonstrated the potential of ML models in long-term delay predictions, it was limited by its focus on individual airports and a reliance on historical datasets from 2011 to 2021, which may not fully capture future or emerging trends in weather patterns. In a similar study, [10] researchers also trained machine learning models with the goal of classifying whether a flight was delayed or on-time due to weather conditions. Their highest accuracy percentage was 80.36% using the Random Forest classifier. Each of these studies focused on a binary classification model in which the aim was to classify a delayed or on-time flight due to various weather conditions without seeking to estimate the duration of the delays. Though several studies have been conducted in which delay classification prediction was explored, the scope of the datasets and the use of quantitative models has been limited.

Building on these foundational studies, our research aims to address the limitations

of previous work by integrating recent, high-frequency data across a more diverse range of U.S. airports. Unlike previous studies that focused on single airports or had limited data, we will use advanced model selection and real-time data integration to improve prediction accuracy and generalizability. In developing models that account for the dynamic interactions between weather conditions and flight delays, we seek to provide stakeholders with actionable insights to reduce delay-related costs, environmental impacts, and improve overall passenger satisfaction. This research will not only advance the current state of delay prediction but also contribute to the broader field of transportation analytics, offering scalable solutions for mitigating weather-related disruptions across various modes of transportation. By using our diverse high-frequency dataset we will be able to use weather data in order to improve upon the accuracy of existing classification models and use a quantitative model that accurately predicts the duration of weather delays.

2 Source

Having a valid and reliable source to find data for a machine learning project is essential. One of our main goals for this project is to determine the effectiveness of weather data as a predictor for flight delays. To carefully find the relationship between weather and flight delays, we needed a dataset that contained both the weather data and the flight data. However, after exploring online resources for such data, we found that this kind of dataset does not exist. The subsequent alternative to finding the data was to find one source for weather data and one for flight data, and then combine them.

We found the Department Of Transportation (DOT) a credible source for flight data. The website allowed us to extract many important variables related to flight information such as flight date, time, airport identification number, departure time, the number of minutes the flight is delayed (difference in minutes between scheduled and actual departure), and delay causes (cause of the delay such weather delay, carrier delay, security delay, and late aircraft delay). However, since delays can occur for multiple reasons, our project will specifically address the impacts of weather-related delays.

Although numerous sources exist to find weather data such as Weather Underground, and NOAA, we found Iowa Environmental Mesonet (IEM) a reliable source for collecting the weather data. We found this source credible for the following two reasons. First, IEM solely focuses on airport weather data, not only in the US but also at airports around the world, which can inform future work by applying the findings of this paper to global datasets. Secondly, IEM extracts the data through the Automated Surface Observing System and according to the IEM website, “ASOS networks are nationally monitored for quality 24 hours per day”. This adds another level of confidence and surety to the validity and reliability of the data. In our study, we included weather data of all the airports from 10 different states including: Florida, Texas, Colorado, New York, Illinois, California, Georgia, New Jersey, Maryland and Nevada between the years of 2014 and 2024.

After conducting a comprehensive literature review, we identified the key concerns and how different researchers approached weather-related flight delays. Almost all of the past literature reviews have been published in the past 5 years which is very relevant to our research. The oldest one goes back to July 2007, which is still relevant today,

but this paper only talks about the environmental impacts of flight delays and not the technical aspects of it. The studies that we have used as our sources have been published in credible journals in the fields of aviation, aerospace, meteorology, and big data which are all essential fields of studies to understand the complex relationship between weather and flight delays. Not only that, most of our sources have been published by the most credible scientific publishers such as Springer, and Cambridge University Press. These interdisciplinary publications provide us with a strong foundation on past work completed in this field and ensure the reliability and validity of our sources.

3 References

[1] Yhdego et al. "Analyzing the Impacts of Inbound Flight Delay Trends on Departure Delays Due to Connection Passengers Using a Hybrid RNN Model." Mathematics 2023, 11, 2427.

[2] Kim, S., Park, E. Prediction of flight departure delays caused by weather conditions adopting data-driven approaches. J Big Data 11, 11 (2024).

[3] Goodman, C. J., and J. D. Small Griswold, 2019: Meteorological Impacts on Commercial Aviation Delays and Cancellations in the Continental United States. J. Appl. Meteor. Climatol., 58, 479–494

[4] Kiliç, K.; Sallan, J.M. Study of Delay Prediction in the US Airport Network. Aerospace 2023, 10, 342

[5] Gratton, G. B. et al. "Reviewing the Impacts of Climate Change on Air Transport Operations." The Aeronautical Journal 126.1295 (2022): 209–221. Web.

[6] Tileagă, C., Oprișan, O. (2021). Flights Delay Compensation 261/2004: A Challenge for Airline Companies?. In: Orăștean, R., Ogrean, C., Mărginean, S.C. (eds) Organizations and Performance in a Complex World. IECS 2019. Springer Proceedings in Business and Economics. Springer, Cham.

[7] Carlier, Sandrine, Ivan de Lépinay, Jean-Claude Hustache, and Frank Jelinek. Environmental Impact of Air Traffic Flow Management Delays. CiteSeerX, doi:10.1.1.76.3545.

[8]Sekelová, I., Korba, P., Pjurová, S., Marimuthu, S., Kale, U. (2023). Reducing the Environmental Impact of Aviation by Minimizing Flight Delays. In: Sogut, M.Z., Karakoc, T.H., Secgin, O., Dalkiran, A. (eds) Proceedings of the 2022 International Symposium on Energy Management and Sustainability . ISEMAS 2022. Springer Proceedings in Energy. Springer, Cham.

[9] Hsu et al. "Unraveling Extreme Weather Impacts on Air Transportation and Passenger Delays using Location-based Data."

[10] "Prediction of Weather-induced Airline Delays Based on Machine Learning Algorithms." 2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC), Digital Avionics Systems Conference (DASC), 2016 IEEE/AIAA 35th.